

## 高臨場感多地点音声コミュニケーションシステムの開発と評価

箕浦 大祐<sup>†</sup> 小長井俊介<sup>†</sup> 森内万知夫<sup>†</sup> 小林 稔<sup>†</sup> 石橋 聡<sup>†</sup>

日本電信電話株式会社 {<sup>†</sup>NTT サイバースペース研究所, <sup>†</sup>NTT サイバーソリューション研究所}

CAVE のような没入型空間と PC をクライアントに有する異機種端末間の高臨場感・多地点音声コミュニケーション環境を実現した。本システムでは、音声ミキシング処理を利用者端末 PC で行うことによって伝送遅延をミキシング人数に依らず約 220 ミリ秒に抑制することができた。音声の再生環境については、5.1 マルチチャンネル音声を利用して立体音場を提供する没入型空間に適した環境。および PC クライアント利用者が利用するヘッドホンによる環境を実現し、その特徴について調べた。本報告では、本システムのアーキテクチャについて述べるとともに、評価実験を通して得られた知見について考察する。

### Development of a Multipoint Realistic Voice Communication System and Its Evaluation

DAISUKE MINOURA<sup>†</sup>, SHUNSUKE KONAGAI<sup>†</sup>, MACHIO MORIUCHI<sup>†</sup>, MINORU KOBAYASHI<sup>†</sup> AND SATOSHI ISHIBASHI<sup>†</sup>

{<sup>†</sup>NTT Cyber Space Laboratories, <sup>†</sup>NTT Cyber Solution Laboratories}, NTT Corporation

We have developed a multipoint realistic voice communication system with heterogeneous client environments. The client environments are an immersion display space and an ordinary PC. Voice communication delay of the system is suppressed in 220-milliseconds approximately, with voice mixing process in a client side. We adopt a 5.1 multi-channel loudspeaker system for the immersion space client and a headphone for the PC client. This report shows implementation and performance of the system, and also discusses characteristics of the system revealed through some evaluation experiments.

#### 1. はじめに

臨場感の高い多人数会話環境にとって、聴覚情報は視覚情報と並んで重要な要素であることが言われてきている<sup>1)</sup>。それにもかかわらず、視覚情報の高臨場感没入型環境については CAVE (CAVE Automatic Virtual Environment<sup>2)</sup>) 間通信のような研究が数多く見られることと比較して、聴覚情報の高臨場感没入型環境については文献 3) にわずかに試みがある程度である。提供される聴覚情報の検討の多くは、会話が成立すればよい程度<sup>4)</sup>か、単体端末での高臨場感環境音の提供<sup>5)</sup>であり、多人数会話没入型環境としては検討が始まったばかりといえよう。

筆者らは、PC と CAVE を利用者端末に含む、高臨場感な会話音声を伝達する多地点コミュニケーション環境を実現した。音声は送信側の PC で符号化処理され、受信側 PC で符号伸長を行った後に、サウンドカードで音声ミキシング処理を行い、立体的に再生する。この方式によって、従来のテレビ会議システムのようなサーバキシングと比較して遅延時間が大幅に短縮した。また、5.1 マルチチャンネルスピーカシステムを利用して没入型空間の装置構成に適した立体音場を提供する端末環境の構築方法も検討した。

以降の章では、システムのアーキテクチャと音声再生環境について述べ、性能評価実験および音源方向定位に関する主観評価実験を通して得られた知見について考察する。

#### 2. 実現したコミュニケーションシステム

##### 2.1 システム構成および CAVE 端末環境

システムの全体構成について図 1 を用いて述べる。没入型端

末は、映像表示用計算機と音声制御 PC を有する。映像表示用計算機には OS に IRIX64 Release 6.5 を搭載した ONYX2 IR2 を用い、ステレオ視による立体表示可能な仮想空間と利用者の分身であるアバタ (avatar) をコンピュータグラフィックスで表示する。音声制御 PC には OS に Windows2000、CPU に Pentium III 850MHz、そしてサウンドカードに Creative Technology 社の Sound Blaster Live! Platinum を搭載した DOS/V PC を用いている。スピーカシステムは、このサウンドカードが推奨する Cambridge SoundWorks 社の DeskTop Theater PlayWorks 2500 を採用した。映像表示用計算機が把握する利用者の視点位置と方向に自分のアバタが配置されて他の利用者から観察される。音声制御 PC はこの視点位置と方向を常に追従し、利用者の視点に合わせた立体音場を生成する。

映像提示装置は、正面、左右、床面の 4 面に 3メートル四方の大きさの背面投射スクリーンを利用している。この装置に合わせたスピーカ配置を検討した。高い臨場感を得るためのスピーカ配置について、国際電気通信連合 ITU の勧告に記述がある<sup>7)</sup>。この勧告を参考に、かつスピーカの影がスクリーンに映りこまないように立体的な音場の実現を考え、左右と正面スクリーンの支柱の背後 4 方向にスピーカを設置した。ところで、スクリーンの背後にスピーカを設置する場合、音圧でスクリーンがゆれる恐れがあるという報告がある<sup>8)</sup>。そこで本システムでは図 2 のように、5.1 マルチチャンネル音声用スピーカシステムからセンタースピーカを除き、スクリーンのゆれの原因になりかねない低周波成分の再生を分離して、サブウーファー (SW) として背面の開口部に設置し、前方左右 (FL, FR) と背後左右 (RL, RR) に高周波再生用のスピーカを設置する方式を採った。SW は床面に置き、それ以外

のスピーカは 1.2メートルの高さに設置した。

PC 端末は、映像表示と音声制御を 1 台の PC で行っており、没入型端末の音声制御 PC と同じ仕様のハードウェアと OS を用いている。没入型端末利用者と同じ仮想空間に参加でき、没入型端末と同じ音声会話機能を利用できる。

サーバは OS に Solaris 5.7 を搭載した SUN Ultra 1 workstation を利用した。

## 2.2 会話音声の処理方式

音声通信機能以外のシステム詳細は文献8)を参照していただくことにし、ここでは音声処理方式について詳しく述べる。

多地点音声コミュニケーションシステムを検討する場合、主な課題にスループットと伝送遅延の関係が挙げられる。従来はインターネットへのダイヤルアップ接続による利用を想定して、図3(a)のようにサーバクライアント方式の構成を採り、サーバでミキシングを行ったのちに配信することでスループットを抑制する方法が多くのシステムで採用されている<sup>9-11)</sup>。サーバでミキシングを行うと、受信する音声データ量は1人の送信者がサーバに送る量と同じで済むが、サーバでデータを伸長処理と再符号化処理を行うため、ミキシングを必要としない1人分のデータ送信に比べて伝送遅延が増す。

会話の臨場感を優先して考える場合、声の遅延は円滑な会話に支障が生じる。そこで今回実現したシステムでは、伝送遅延の抑制を優先し、図3(b)に示すように送信された音声データは受信側端末でミキシングさせる設計とした。受信側はサーバとの間に発話者の数だけデータ一時蓄積用のバッファを設ける。ただし、受信側計算機で管理可能なバッファ数はその処理性能に依存するので、作成可能なバッファ数を自動的に検知し、会話空間への参加時にあらかじめサーバに通知しておく。サーバは、このバッファ数の設定に基づいて受信可能な人数分の音声データを配送制御する<sup>12)</sup>。

ネットワーク環境は、現在の LAN や近い将来の光 IP ネットワークを想定して 10Mbps LAN を利用しているが、利用者数の増加に比例して音声ストリーム数が増加する上に、音声以外のコミュニケーションデータも伝送したいと考えるとデータ圧縮が必要である。符号化方式については、没入型空間での楽器演奏などの会話以外の楽音通信にもある程度対応し、かつアルゴリズム的に遅延が少ない方式の検討を念頭におき、サンプリング周波数 22.05kHz、量子化ビット数 16 ビットの ADPCM を採用することで、良好な音質とビットレート抑制のバランス調整を図った。

会話音声の処理方法について図4を用いて述べる。マイクを通して PC に取り込まれる音声は、ADPCM による符号化を行った後、利用者識別子とともにサーバに送信される。利用者は音声送信のオプションとして、常時音声入力受付モード(Auto Voice Sending, 以下 AVS)と Shift キー押し下げ時音声入力受付モード(Push To Talk, 以下 PTT)を選択できる。これは無音データの無駄な送信を避けるための機能である。AVS では一定音量以下の

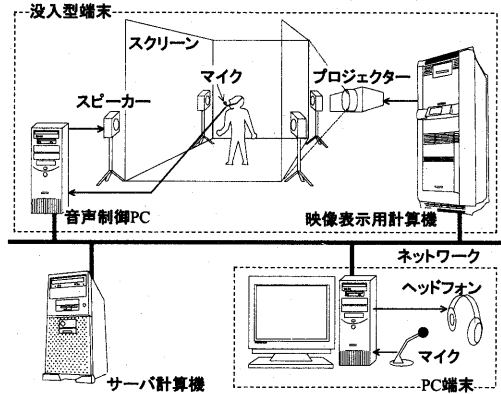


図1 システム構成  
Fig. 1: System components.

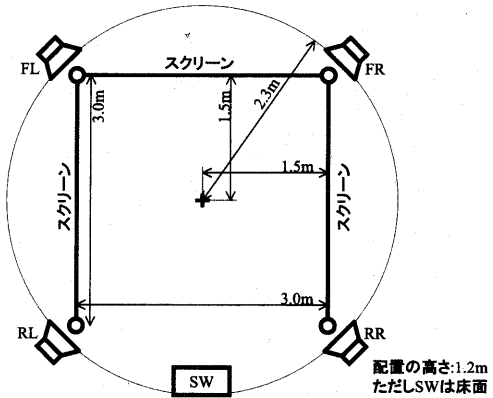


図2 没入型端末におけるスピーカ配置  
Fig. 2: Loudspeaker arrangement for an immersion display client.

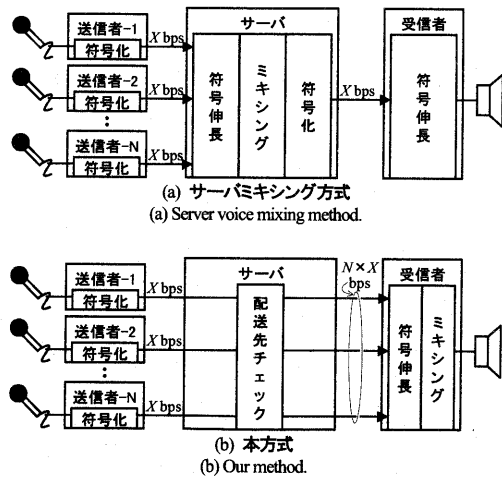


図3 デジタル音声データの伝送とミキシング方法  
Fig. 3: Digital sound communicating and mixing methods.

場合はデータ送信を行わない無音検出機能を実装したが、利用環境が雑音の多い場所を想定すると PTT も必要となる。PTT は無音検出を行わない。すなわち、Shift キーが押されている間は無音の空データとして送信される。

他の利用者から音声データを受信した場合は、符号伸長を行った後、利用者識別子別にバッファに振り分ける。振り分けは、すでに同一の識別子を有するデータを蓄積するバッファが存在すればそのバッファに、新規の識別子の場合は空きバッファを探してデータを渡す。作成可能なバッファ数は、最大 16 個まで可能な設計としたが、前節で述べた性能の PC の場合、自動的に 8 個までの設定となる。

ミキシング処理および立体音場生成は Windows2000 の DirectSound の機能を使って実現される。DirectSound を使うことで、音源の距離と方向に応じた音圧レベルとバランス制御が可能となる。また、DirectSound 対応のサウンドカードを使うことで、ミキシングと立体音場の生成はサウンドカードで行い、PC の計算負荷を軽減するアーキテクチャとした。したがって PC の CPU における音声再生に関する処理は、受信音声バケットの振り分けと符号の伸長処理のみで済む。

会話音声以外のコンテンツとして使われる効果音は、仮想空間内の音源位置と利用者の観察視点の位置と方向をもとに、各端末で独立に DirectSound によって立体的に再生される。

### 3. 再生される会話音声の特徴および考察

#### 3.1 会話音声の伝送性能

本システムの性能を明らかにするために発話から再生までの伝送遅延の測定と、ネットワークトラフィックの測定を行った。以下の実験では、図5に示すサーバ1台と没入型端末2台、PC端末7台のシステムに加え、入力音声生成用PC、遅延測定用PC、ネットワークトラフィック測定用PCによる装置構成とした。入力音声生成用PCの再生音声は、1台のPC端末のマイク入力(MIC-IN)と遅延測定用PCの音声左入力(L-IN)に分岐して入力される。また、別のPC端末のスピーカ出力(SPK-OUT)を遅延測定用PCの音声右入力(R-IN)に接続した。入力音声生成用PCにあらかじめWAVE形式の音声ファイルを用意した。この音声ファイルは図6(a)に示すように3秒間隔でホワイトノイズが記録されている。この音声を再生すると、遅延測定用PCには図6(b)のようにL-INには直接話者側の音声、R-INにはシステムを通じた聞き手側で再生される音声が入力される。このPCで音声データ編集ソフトを用いて入力音声データを左右別に取得し、有音部の立ち上がりの差を遅延として測定した。音声データ編集ソフトは入力音声を44.1kHzのサンプル単位で解析可能である。全端末の音声入力モードはPTTとし、入力音声生成用PCから音声入力されない端末のマイク入力に何も接続しない。発話状態を作り出す時は、Shiftキーにおもりを乗せることで、常に無音データがサーバに送られ、受信側ではミキシング処理される。

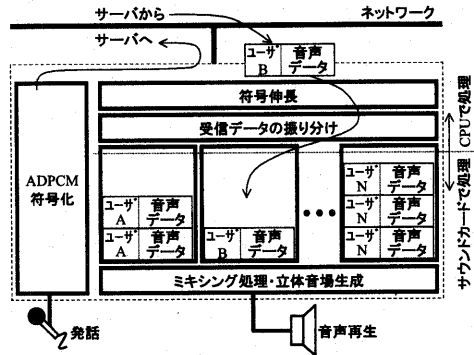


図4 端末の音声処理  
Fig. 4: Voice data processing in a client PC.

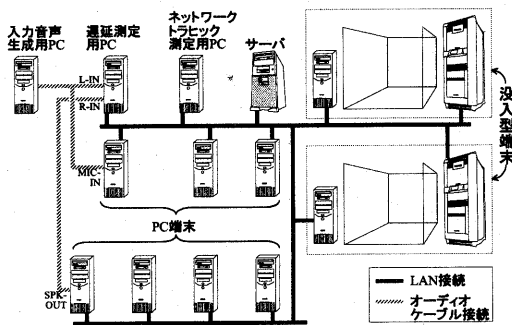
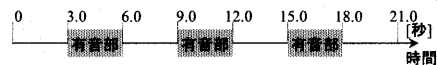
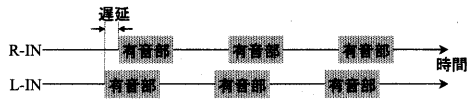


図5 実験装置構成  
Fig. 5: Experimental equipment arrangement.



(a) 入力音声生成用PCの出力音声  
(a) Output sound from the input voice generating PC.



(b) 遅延測定用PCの入力音声  
(b) Input sound to the delay measuring PC

図6 実験で使った音声波形  
Fig. 6: Experimental sound waveform.

ネットワークトラフィック測定は、ネットワークトラフィック測定用PCで測定ソフトウェアを用いて特定の1台の端末がサーバから受信するトラフィック量を測定した。

比較のために、Microsoft社のNetMeetingと、InterSpaceシステム<sup>13)</sup>を同じシステム構成で稼働させて音声の伝送遅延とトラフィック測定を行った。NetMeetingはPCM 64Kbpsを、InterSpaceはGSM 06.10を採用している。NetMeetingは、First Virtual Communications社のMeetingPointを多地点会議サーバとして使用することで、最大8音声のミキシングが可能となるので、これを

利用した。

図7に遅延測定の結果を示す。横軸は発話者の数を、縦軸が遅延時間を示す。InterSpaceについては、ミキシング可能上限の2人まで<sup>14)</sup>の測定である。各測定値は、30回の試行、すなわち図6(a)の入力音声の再生1回で測定できる3試行分を10回行った平均値である。結果を見ると、本システムの遅延は、比較システムの半分以下の220ミリ秒程度である上に、人数が増えてもほぼ一定の遅延を保っている。誤差を比較すると、遅延のゆらぎも低い。ミキシング人数が増加しても遅延が一定であることは、安定した会話環境を提供する上で重要である。ITUではリアルタイム通信の許容範囲を400ミリ秒までと勧告しており<sup>15)</sup>、十分満足できる性能と言える。

一方で、全く気にならない遅延の許容範囲となるとさらに抑制を求められる。上記の勧告でも150ミリ秒以下が最もよいと述べており、ある報告では、許容遅延は往復で100ミリ秒までであり、別の報告にも遅延が100ミリ秒以上になると通常の会話に影響が出る<sup>16)</sup>。これら報告を参考にすれば、遅延は50~150ミリ秒以内に抑えることが望ましいが、図8のようにミキシング人数に比例して受信データ量が増えるので注意が必要である。

### 3.2 端末環境の音声再生特性

前節で明らかになった処理性能の音声コミュニケーションを利用する端末での音場特性について調べた。実験はスクリーン内空間の音圧レベル分布と音源からの距離に応じた音圧レベルの減衰、そして没入型端末とPC端末のヘッドホン双方で音源方向の定位の正確さについて行った。会話での利用を想定しているため、実験対象は利用者の耳の高さの水平面内とする。

実験の前に、スピーカ音圧レベルの前後バランス調整を行った。前後のスピーカと利用者の間には、スクリーンなどの遮蔽物が非対称に存在するため、音圧レベルの減衰量の予想が難しい。そこでup-and-down法<sup>17)</sup>による前後スピーカ出力バランスの調節を行い、感覚量として等しい設定を探した。調整と合わせて音が聞こえる時にスクリーンのゆれを感じるか聞き取りを行ったが、感じると答えた被験者はいなかった。

この調整を経て、音源がスクリーン内の中心にあるとしてスクリーン内の音圧レベル分布を測定した結果が図9である。1kHzの純音を30秒間再生し、その平均音圧レベルを25cm間隔で測定した。測定の高さは1.6メートルである。グラフはスクリーン内中心を原点とし、右側スクリーン(RIGHT)をX軸正の方向、正面スクリーン(FRONT)をY軸正の方向として等音圧レベル線で表されている。座標(150, 150)での音圧レベルは87dBだが、座標(0, 0)に届く時には約70dBに減衰する。利用者の背面(REAR)はスクリーンのない開口部であるため若干特性が異なる。開口部の減衰が少ないためか、座標(150, -150)では85dBであり、後部スピーカの出力は前部より若干低めの設定となる。ただし、スピーカFRとRRの目の前では各93dB、89dBであり、スクリーンなどの遮蔽物による減衰が大きい。

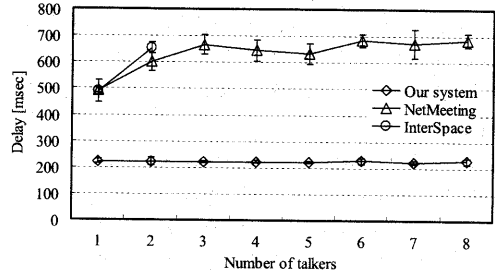


図7 発話者数と伝送遅延の関係

Fig. 7: Communication delay for the number of talkers.

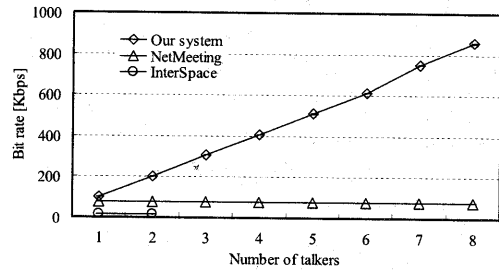


図8 聞き手側の端末1台がサーバから受信するトラフィック量

Fig. 8: Network traffic of one listener client PC receiving from a server.

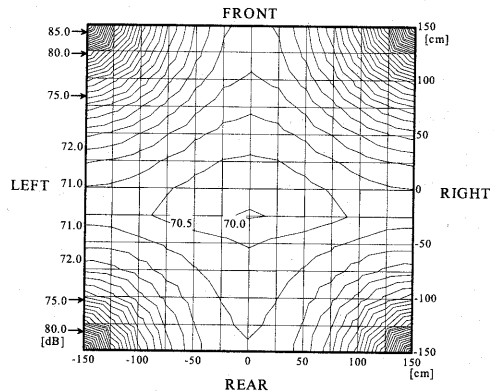


図9 スクリーン内の音圧レベル分布

Fig. 9: Sound pressure level distribution in an immersion display.

音源の距離定位については、空間の反響など経験的知識に大きく左右されることが古くから言われており、特に無響室では正面の音源の距離を全く半別できない<sup>18)</sup>。最近の研究でも、聞きなれない音源の距離知覚の誤差が20%台であっても不思議ではないといわれている<sup>19)</sup>。今回開発したシステムも特殊な音響空間と考えられる。また、表示するコンテンツの空間構成に依存して、視覚的刺激から想像される再生音の音圧レベルと音源までの距離の関係も変化すると考えられる。そこでまず、音の空間伝播における距離減衰の特性をどの程度再現できるか調べた。

図10に音源からの距離減衰を測定した結果を示す。測定はスクリーンの中心、座標(0, 0)の位置の高さ1.6メートルで行い、こ

の位置で利用者が正面スクリーンに正対した時に、正対する音源までの仮想空間内の距離を変化させた。音源は音圧レベル分布の測定時と同じ純音を使い、30秒間再生した時の平均音圧レベルとして測定した。グラフの横軸は音源からの距離、縦軸が音圧レベルを示す。また、◇は測定値、細い実線はシステムが期待する出力値の設定である。理論的には音圧レベルは距離の自乗に比例して減衰することを踏まえて、グラフに理想値として破線で示した。しかし、DirectSoundでは距離に比例して音圧レベルが減衰するように設定されている。したがってグラフの細い実線は破線と一致しない。また、不要な距離計算を避けるため、設定した範囲内で距離減衰が生じるように設定可能なので、音源までの距離が1メートルより近くても1メートルの時と同じ音圧レベル、また、10メートル以上離れても10メートルの時と同じ音圧レベルとして設定している。測定値を見ると、スクリーンなどの障害物による減衰のためか、システムとして期待した出力値よりも低い値を示している。現実空間での特性を厳密に再現しているとは言いがたいが、それに近い形で距離に応じた減衰制御が反映されていることがわかる。

次に、音源定位の実験について述べる。実験は、被験者を囲むように仮想空間内の2メートル離れた円周上に男性の容姿のアバタを立て、アバタのいずれかの位置から男性の声で「こんにちは、久しぶりだね」と声をかける状況を設定して行った。音源方向は30度間隔で、アバタは音源がとらうる位置に計12体配置した。没入型環境では、被験者は立体視用の眼鏡をかけて、どのアバタが発声したと思うかを答える。PC端末ではヘッドホンを装着して、仮想空間が表示される15インチのディスプレイを見ながら回答する。アバタの配置を30度ごとにした理由は、人間の音源定位の精度を参考にしている。音源が正面の時は定位の誤差は5度程度、正面から30度から90度離れた方向では誤差が15度程度であると言う<sup>20)</sup>。また、音だけを手がかりにした音源定位が若干ずれても、腹話術効果<sup>21)</sup>によって視覚的な手がかりに誘導されて、結果的に正しく定位できることが十分期待できる。本システムは総合的に感覚を利用するコミュニケーションシステムを目指しており、このような相乗効果は積極的に利用したい。そのため、前述の報告を参考に±15度の範囲で音源の方向を定位できることが確認できれば十分であろうと考えた。

図11に実験結果を示す。グラフの各軸は音源方向を表す。試行は8人の被験者に対して各方向5回、ランダムな順序で行った試行結果の正解率を表示している。実線は回答が音源方向と完全に一致した場合、点線は正解方向±30度も正解に含めた場合である。

没入型端末の結果をみると、-120~120度の範囲では50%以上の確率で良好に音源定位がされている。±30度の誤差も含めれば、80%以上の正答率を示している。ところが、±150度と180度については正答率が低い。原因として、5.1マルチチャンネルのスピーカシステムでも背後の音源定位は前方と比較して難し

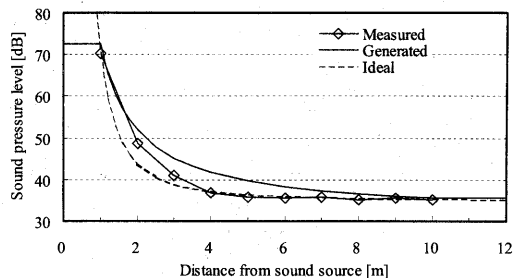
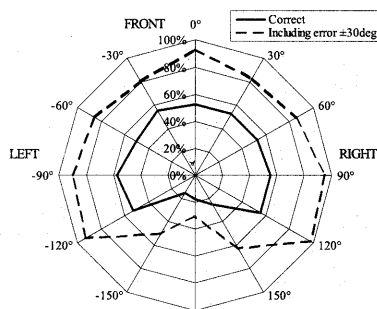
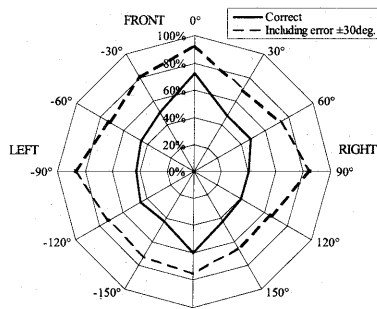


図10 音源からの距離による音圧レベルの減衰

Fig. 10: Sound pressure level attenuation for distance from sound source.



(a) 没入型端末での結果  
(a) For an immersion display client.

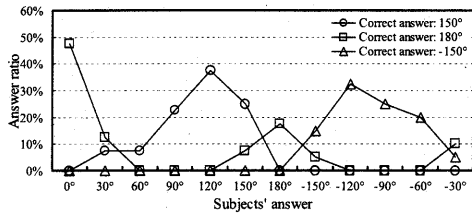


(b) PC 端末での結果  
(b) For a PC client.

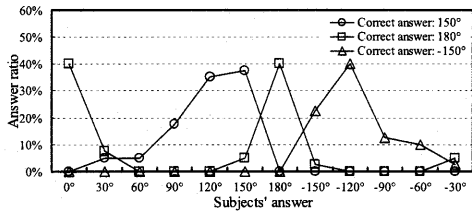
図11 音源方向定位の精度

Fig. 11: Accuracy of sound azimuth localization.

い<sup>22)</sup>ことや、背後はスクリーンがないので視覚的な手がかりがないことなどが推測されるが、本来音源定位は、利用者の正面と後ろの正面の精度は高い<sup>23)</sup>はずである。ここで、文献24)に興味深い報告がある。音源が被験者の真後ろ(本実験でいう180度)に設置した場合、音声に合わせた話者の映像が真正面(本実験でいう0度)に提示されれば、300秒間の刺激提示に対して約260秒間の腹話術効果が生じるという。そこで、±150度と180度音源方向に対する回答の内訳を図12(a)に示した。正解が±150度の場合は、回答の誤差が大きく定位の精度の問題が強いと思われる



(a) 映像表示あり  
(a) With visual cues.



(b) 映像表示なし  
(b) Without visual cues.

図 12 没入型環境での音源が±150度、180度の場合の回答内容  
Fig. 12: Answers for sound azimuth ±150° and 180° in an immersion display.

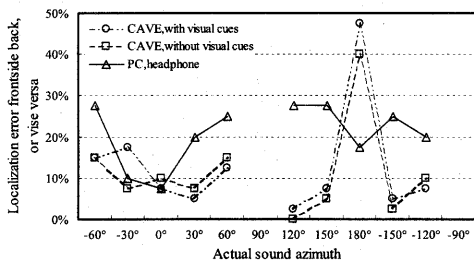


図 13 前後の対称方向に定位判断を誤った割合  
Fig. 13: Error rate of localization front side back, or vise versa.

るが、正解が 180 度の場合には、確かに 0 度と回答するケースが多い。したがって、上記の原因に加えて視覚刺激による定位の誘導も原因に考えられる。試みにアバタを表示せずに同じ条件、同じ被験者に音源定位をさせた(図 12(b))。正答率は向上したが、依然として同様の傾向があり、視覚刺激だけが主原因とは言いつれない。

ヘッドホン条件では没入型環境と異なり、方向定位は正面と真後ろでは良好だが左右の方向では劣っている。回答の内容を調べると、図 13 のように±90 度の直線に対称な判断の誤りが、没入型環境よりヘッドホン条件のほうに多いことがわかる。このような前後の対称方向についての判断の誤りは非常に多いことが報告されている<sup>25)</sup>。没入型環境では刺激提示中にも頭を自由に動かして音源方向を確認する動作を行うことが可能であるが、ヘッドホンで刺激を提示される場合にはこれはできないために誤りが多くなっていると推察される。

#### 4. おわりに

会話などの入力音声データを PC で処理し、計算機ネットワークを通して伝送する没入型環境のための多地点音声コミュニケーション環境を実現した。音声データのミキシング処理を利用者端末側で行うことによって、従来のテレビ会議システムなどが採用するサーバミキシング方式と比較して音声の遅延を大幅に抑えることができた。また、没入型ディスプレイを用いた端末に 5.1 マルチチャンネル音声を利用した立体音場の再生環境を構築することで視聴覚両面から臨場感の高い会話環境を実現した。

しかし、音声遅延の許容範囲や聞き手側端末の音場再現は、現実空間の対面コミュニケーションのレベルには達していない。視覚・聴覚の両モダリティを総合的に利用する実際の会話状況を想定した利用実験を通して、さらに臨場感を高めるためのコミュニケーション要素や技術的課題の抽出を行ってきたい。

#### 参考文献

- 1) 例えば, C. Hendrix, et al.: "The sense of presence within auditory virtual environments," *Presence*, Vol.5, No.3, 1996.
- 2) C. Cruz-Neira, et al.: "Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE," *SIGGRAPH 93*, 1993.
- 3) J. Pair, et al.: "COOLVR: Implementing audio in a virtual environments toolkit," *ICAD97*, 1997.
- 4) C. Greenhalgh, et al.: "Making networked virtual environments work," *Presence*, Vol.10, No.2, 2001.
- 5) 土田ほか: "音響シミュレーションに基づいた CABIN における 3 次元音響提示," *VR 学大*, 1999.
- 6) 久木元ほか: "プロジェクション型没入ディスプレイにおける 3 次元音場生成のための実用的検討," *信学技報 EA99-32*, 1999.
- 7) ITU-R BS.775-1: "Multichannel stereophonic sound system with and without accompanying picture," 1992-94.
- 8) 箕浦ほか: "千人規模の利用者のための 3 次元仮想空間コミュニケーション環境," *情報学論*, Vol.42, No.11, 2001.
- 9) ITU-T H.323: "Packet-Based Multimedia Communications Systems," 2000.
- 10) S. E. Chodrow, et al.: "Design and Implementation of a Multicast Audio Conferencing Tool for a Collaborative Computing Framework," *3rd Joint Conf. on Information Sciences*, 1997.
- 11) B. Damer: "Inside Traveler's Voice Codec," *Avatars!: Exploring and Building Virtual Worlds on the Internet 付属 CD*, Addison-Wesley, 1997.
- 12) 小長井ほか: "共有仮想空間における音声コミュニケーション方式の検討," *信学技報 MVE99-73*, 2000.
- 13) S. Sugawara, et al.: "InterSpace: Networked Virtual World for Visual Communication," *IEICE Trans. Vol.E77-D*, No.12, 1994.
- 14) 大井ほか: "インタースペースにおける音声制御方式," *NTT R&D*, Vol.47, No.4, 1998.
- 15) ITU-T G.114: "One-way transmission time," 2000.
- 16) E. M. Schooler, "Conferencing and Collaborative Computing," *Multimedia Systems*, Vol.4, No.5, 1996.
- 17) 日科技連官能検査委員会編: *新版官能検査ハンドブック*, 日科技連出版社, 1973.
- 18) 境ほか: *聴覚と音響心理*, コロナ社, 1978.
- 19) B. C. J. Moore: *An introduction to the psychology of hearing*, Academic Press, 1997.
- 20) S. S. Stevens, et al.: "The localization of actual sources of sound," *J. Experimental Psychology*, Vol.48, 1936.
- 21) 例えば, W. R. Thurlow, et al.: "Certain determinants of the 'ventriloquism effect,'" *Perceptual and Motor Skills*, Vol.36, 1973.
- 22) 濱崎: "マルチチャンネル音声収録とデジタル放送," *音響誌*, Vol.57, No.9, 2001.
- 23) (社)電子通信学会編: *聴覚と音声*, コロナ社, 1966.
- 24) C. E. Jack, et al.: "Effects of degree of visual association and angle of displacement on the 'ventriloquism' effect," *Perceptual and Motor Skills*, Vol.37, 1973.
- 25) 中林: "水平面内における方向定位," *音響誌*, Vol.30, No.3, 1974.